DOCUMENT RESUME

ED 143 706 TM 006 495

AUTHOR Klein, Anna C.; Whitney, Douglas R.
TITLE A Discussion and Comparison of Selected Methods for Determining Cutoff Scores for Proficiency and Placement Tests. Placement and Proficiency Testing Report No. 6.
INSTITUTION Iowa Univ., Iowa City. Evaluation and Examination Service.
PUB DATE Apr 77
NOTE 60p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS *College Placement; *Cutting Scores; Decision Making; *Equivalency Tests; Evaluation Methods; Higher Education; Institutional Research; *Mathematical Models; National Norms; Norms; Test Results; *Tests
IDENTIFIERS College Level Examination Program; *Trait Treatment Interaction

ABSTRACT

Procedures and related issues involved in the application of trait-treatment interaction (TTI) to institutional research, in general, and to placement and proficiency testing, in particular, are discussed and illustrated. Traditional methods for choosing cut-off scores are compared and proposals for evaluating the results in the TTI framework are presented. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED143706

A DISCUSSION AND COMPARISON OF SELECTED METHODS FOR DETERMINING CUTOFF SCORES FOR PROFICIENCY AND PLACEMENT TESTS

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Anna C. Klein
Douglas R. Whitney

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Anna C. Klein

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NIE SYSTEM.

Placement and Proficiency Testing
Report No. 6

A paper presented at the annual meeting of the American Educational Research Association
New York
April, 1977

Evaluation and Examination Service
University of Iowa
Iowa City, Iowa

April, 1977

IM006 495

ABSTRACT

Choosing cutoff scores for proficiency and placement tests is set into the framework of trait-treatment interactions (TTI). Methods for choosing cutoff scores are discussed in the context of amount and type of data available and comparisons are made on that basis. Two TTI methods for choosing or evaluating cutoff scores are proposed.

ORGANIZATION

Page

PUTTING PLACEMENT AND EXEMPTION INTO THE TTI FRAMEWORK 1

Frame of Reference. 1

Trait-Treatment Interactions. 1

Major Classes of Alternate Treatments in Education 7

Models for Placement and Exemption Testing. 11

Application of TTI to Placement and Exemption Models. 12

METHODS FOR DETERMINING CUTOFF SCORES. 14

No Empirical Data 15

External Norm Data. 16

Choosing Cutoff Score Based on National Norms. 16

Minimum Recommended Cutoff Scores Based on Chance Performance. 18

General Strategies for Placement and Exemption 20

Local Norm and Criterion Data 21

Local Norm Data Without Criterion Data 21

Discriminant Analysis Procedure. 23

Equal Percentile Method. 27

Regression Method. 27

Relative Seriousness of Classification Errors 30

Other Related Measures. 34

To Establish the Decision Rule 34

To Explain a Result. 35

For Future Counselling 38

Related Issues. 39

Motivation 39

Sample Size. 41

Grades 41

PROPOSED MODELS FOR EVALUATING CUTOFF SCORES USING TTI 43

Placement 43

Exemption 46

LIST OF TABLES

Table		Page
1	Major classes of alternate treatments.	8
2	Distinction between the character and use of placement and exemption tests.	9
3	Minimum cutoff scores based on the mean and standard deviation of chance score.	19
4	Scores of University of Iowa students on the CLEP Educational Psychology subject examination	22
5	Percent of students correctly classified using discriminant analysis of MLA Spanish Level L.	25
6	Means and standard deviations obtained for local administration of MLA Spanish Level L	26
7	Decision theory model applied to the validation sample	32
8	Table of estimated proportions for cells of Table 7 based on Educational Psychology students at the University of Iowa .	32
9	Expected losses resulting from using selected cutoff scores for awarding exemption to Educational Psychology students. . .	34
10	CLEP German Pre-Test results for second year students broken down by program choice and reason for taking German.	37
11	Expectancy of Success on the CLEP Introductory Sociology Test Predicted by the ACT Social Studies Score	38

LIST OF FIGURES

Figure		Page
1	No interaction between trait and treatment.	3
2	Disordinal interaction between trait and treatment.	4
3	Ordinal interaction between trait and treatment	5
4	TTI for placement models.	12
5	TTI for exemption models.	13
6	Discriminating between two groups using a test with two subscales	24
7	Regression lines for choosing cutoff scores	29
8	Applying the methods for choosing a cutoff score for CLEP Introductory Sociology based on course performance.	31
9	Random placement of students with trait scores in specified interval.	44
10	Evaluation of a cutoff score that has been used in placement model	45
11	TTI for exemption or credit by examination.	46
12	A proposed method for establishing TTI for exemption studies. .	47
13	Applying the proposed method for establishing TTI to CLEP Introductory Sociology data	49
14	Applying the proposed method for establishing TTI to CLEP Introductory Sociology data after adjustment to educational benefit	50

LIST OF RECOMMENDATIONS

Recommendation		Page
1	The preferred method for evaluating treatments and cutoff scores (decision rules) for assigning students to one treatment or the other is TTI.	7
2	To assure content validity for a placement test, if several instructors teach the course under consideration, each should independently review the test item by item to determine whether the content and skills covered in the item are covered in his/her version of the course.	10
3	Cutoff scores determined using only test content (such as Nedelsky's MPL) should be based on the opinion of several subject matter experts.	16
4	A cutoff score for placement or exemption should be at least two standard deviations above the exepcted chance score.	20
5	Placement decisions should be based on local data and not solely on national norm data.	21
6	Cutoff scores for exemption <u>may</u> be based on national norm data and local content validation.	21
7	We recommend gathering additional information from students whenever possible for the purpose of explaining ambiguous results. This information may include: number of semesters of previous coursework in the area, reason for taking the course, whether or not they took another specific course and grade in that course, what grade they feel they deserve in the course in which the test is being administered.	39
8	The problem of student motivation needs serious attention in planning local validation administration of a test to be used for placement and exemption	41
9	A measure of course performance other than final course grade should be used as a criterion for test validation and establishment of cutoff scores.	42

PUTTING PLACEMENT AND EXEMPTION INTO THE TTI FRAMEWORK

The purpose of this paper is to stimulate research in the application of trait-treatment interaction (TTI) theory to institutional research in general and to placement and proficiency testing in particular. In this paper we discuss and illustrate procedures and related issues involved in such applications. We also compare traditional methods for choosing cut-off scores and make some proposals for evaluating the results in the TTI framework.

Post-secondary educational institutions have become increasingly interested in making accommodations for individual differences for three reasons: increased diversity of college students, emphasis on improving the relevance of education, and economic problems facing universities and colleges; all three prompting the institutions to adapt to a wider variety of students. These accommodations are made in several ways, for example: institutional choice (c[illegible]unity college, vocational school, junior college), program choice within the institution, elective options within the program, alternate treatments within the program, and individualization. In this paper we will restrict attention to alternate treatments within the program choice.

FRAME OF REFERENCE

Trait-Treatment Interactions (TTI). Decision theory is concerned primarily with determining the effect of alternate actions. In placement, the alternate actions are placing the student into first or second semester

of the sequence; in exemption the alternate actions are exempting the student or not exempting him/her from the course. Thus it seems that decision theory has a role to play in placement and exemption techniques.

> After examining placement procedures that are being used and the problems associated with them, after discovering that there is very little published literature describing successful (or even unsuccessful) placement procedures and their associated curricular components, and after becoming acquainted with some of the problems of setting up operational programs based on the classical approaches to placement, one develops a feeling that what is being done is not very sound or successful and that there must be a better way. One way that has not yet received much more than a theoretical or logical exploration, but that seems promising at this point, is the use of the decision theory model (Cronbach & Glesser, 1965, especially chap. 5). (Hills, 1971, p. 712)

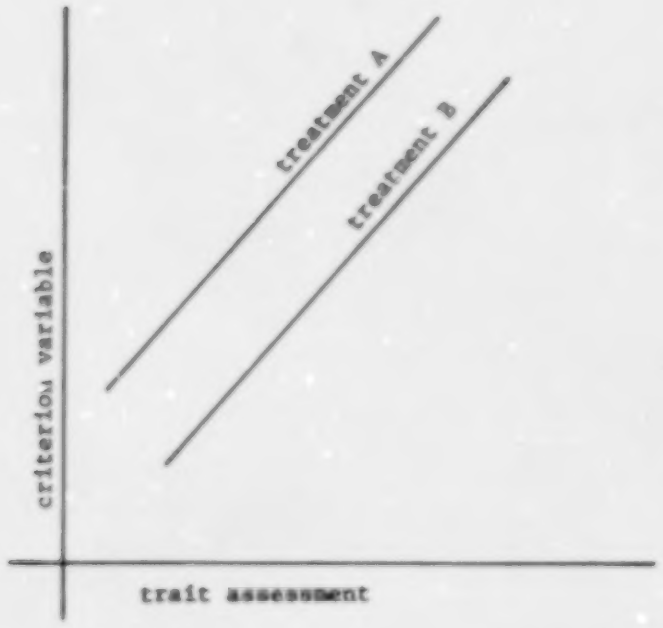
Cronbach is credited with applying the principles of decision theory to alternate treatments in education. The first edition of Cronbach and Glesser's book on such applications was published in 1957, but it seems to have had little effect on the methods used in practice. "It was not until 14 years later that Hills (1971) made the only serious attempt to discuss college placement in the context of decision theory--and that discussion was designed for measurement specialist (see Webb 1967 for a brief exception)." (Willingham, 1974, pp. 10-11).

Willingham (1974) highlights four critical ideas to keep in mind when applying decision theory to evaluating alternate treatments.

1. The procedures always represent alternate treatments.
2. The special treatment is intended only for a proportion of the total population.
3. All consequences (cost, etc.) must be included in the evaluation.
4. The outcome is uncertain when the decision is made.

Suppose that there are two alternate treatments A and B, and two measures are available, one a trait measure for determining which treatment is to be assigned and the other a criterion measure, which are common to the two treatments. Separate regression lines for each treatment are determined from the two measures. Figures 1, 2, and 3 illustrate three possible results of placing students randomly into the two treatments. The outcome in Figure 1 indicates that the trait measure is not differentiating between the two treatments, and thus is not effective for determining placement into treatment A or treatment B. If the

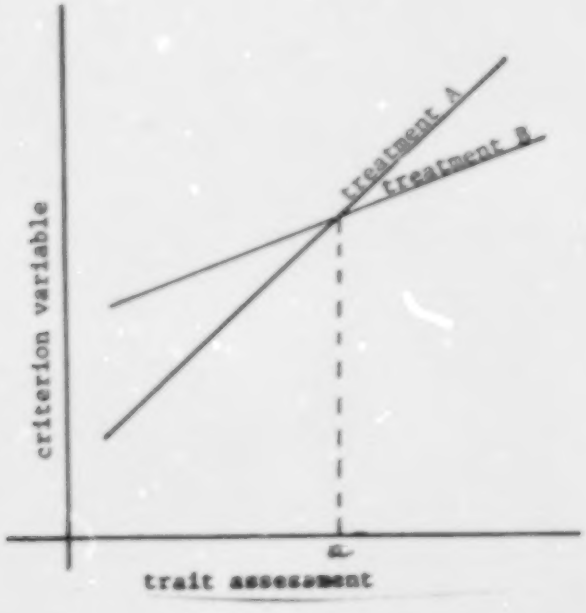
Figure 1: No interaction between trait and treatment.



difference between the parallel lines is deemed significant (taking all relevant cost and benefit variables into account), all students should

be placed into treatment A; if not, either treatment could be used for all students. Figure 2 represents a different situation. In this case the two regression lines have different slopes and intersect within the range of the trait scores. This is called disordinal interaction. This result indicates that students who do poorly on the trait assessment earn a higher value on the criterion variable after the treatment if they are placed into treatment B, whereas, the students who score high on the trait assessment earn a higher value on the criterion measure if they are placed in treatment A. In this situation the best action is

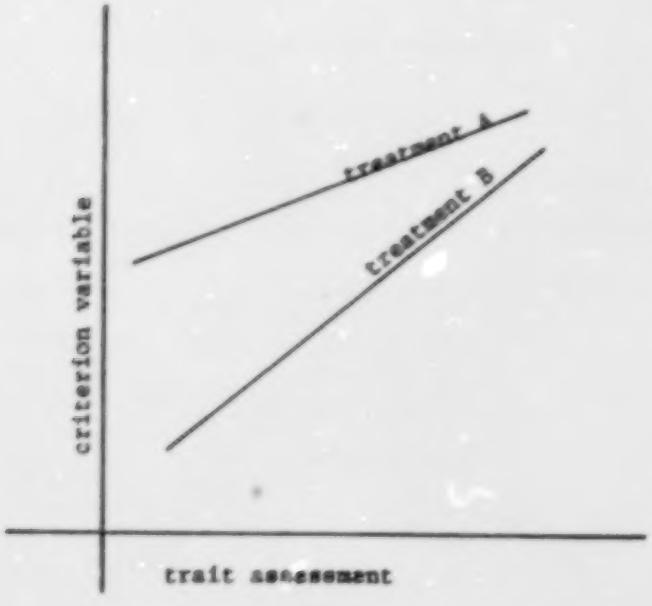
Figure 2: Disordinal interaction between trait and treatment.



clear; students with trait scores below the point a should be placed into treatment B and those with scores above the point a into

treatment A. Some may argue that this is not a very likely situation, and that, in practice, the outcome is more like that presented in Figure 3. The result in Figure 2 could occur if the students who scored high on the trait got bored with the subject matter area as a result of being placed into treatment B, developed poor study habits as a result of material which they already knew, or perhaps were confused by "redundant" presentation.

Figure 3: Ordinal interaction between trait and treatment.



In the possible outcome illustrated in Figure 3, the regression lines have different slopes, but they intersect outside of the range of trait scores. This is called an ordinal interaction. If a final achievement test score were the only variable to be considered and the outcome in Figure 3 was obtained, one would place all students

into treatment A. However, when total utility values are included in the criterion variable (such as time, cost, and other factors), rather than simply the final test score, the regression lines associated with treatments A and B will be adjusted up or down depending upon whether the net utility is positive or negative. These cost-benefit factors will vary for different groups of students, as for example, majors and non-majors in the subject area. "Any dependably established ordinal interaction usually implies a disordinal interaction for some other type of subject." (Cronbach & Snow 1969, p. 19) Thus, ordinal interactions are not to be dismissed lightly and should be considered to be interactions when searching for interactions in educational research. When all relevant factors are considered, the resulting figure may be similar to that presented in Figure 2.

The interaction in either Figure 2 or 3 is called a trait-treatment interaction (TTI) or ATI by some authors for aptitude-treatment interaction. Frequently TTI is used in evaluating the treatments themselves, that is, whether one treatment is better than another for certain students. The intersection point in Figure 2 also provides a basis for a cutoff (decision rule) for placing students into one treatment or the other. Students would be placed into that treatment which results in the highest expected criterion value. So, in the case of Figure 2, students who obtain trait scores above the point a are administered treatment A and the others treatment B.

The above approach using TTI may be contrasted with two other

methods for approaching the study of educational treatments which have been used namely, experimental and correlational approaches. From the experimental point of view, in which only treatment means are considered, treatment A would be chosen for all students in Figures 1 and 3 and probably no significant treatment difference would be detected in Figure 2. Thus, this point of view results in the same conclusion as does TTI only in the case of Figure 1. From the correlational point of view, a high correlation between trait and learning outcome could lead to the conclusion that the trait (placement test) was useful for placement in the case of Figure 1. Thus, each point of view leads to what seems to us to be a faulty conclusion in at least one of the cases presented in Figures 1, 2, and 3, and thus these approaches are inadequate.

> RECOMMENDATION 1: The preferred method for evaluating treatments and cutoff scores (decision rules) for assigning students to one treatment or the other is TTI.

Major Classes of Alternate Treatments in Education. From Figures 1, 2, and 3, we see that three factors are involved in using the TTI approach to evaluating alternate treatments. These are an assessment variable for determining placement, a criterion variable as a common measure of the outcome, and two or more different treatments. To clarify the settings in which the TTI approach could be applied, the trait assessment and the common criterion can be separated into two or more categories. An example of these is summarized in major classes of alternate treatments in Table 1. (Willingham, 1974, p. 18)

Table 1: Major classes of alternate treatments.

NATURE OF TRAIT ASSESSMENT	NATURE OF THE COMMON CRITERION: End of course achievement	NATURE OF THE COMMON CRITERION: Other educational outcomes (persistence, satisfaction, etc.)
Aptitude or personal characteristic	ASSIGNMENT	SELECTION
Knowledge of subject matter	PLACEMENT	EXEMPTION

These four classes are defined as follows.

> Assignment. "Grouping students of similar ability or personal characteristics into alternate educational treatments that have common subject-matter objectives."
>
> Placement. "Positioning students at the optimal point in an instructional sequence on the basis of how much the student knows about the subject."
>
> Selection. "Grouping students of different ability into alternate programs with different educational content."
>
> Exemption. "Excusing students from a degree requirement on the basis of demonstrated proficiency that may have been acquired under any auspices."

In this paper we restrict our attention to the two classes of placement and exemption. As noted in Table 1, the trait assessment for both placement and exemption is an achievement measure. Distinguishing characteristics of the two classes as summarized by Willingham (1974, p. 97) are presented in Table 2. (We use the terms exemption and proficiency interchangeably in this paper.)

The three components of two or more treatments, the variable for determining placement, and the criterion variable have been recognized.

In the case of placement, we will be concerned with treatments which vary according to level or pace of subject matter presented; the placement variable will be a measure of student achievement in the subject; and a measure of achievement may serve as the common criterion for evaluating performance at the end of the sequence. We, however, propose using a measure of educational benefit which incorporates cost-benefit factors, such as time, as the criterion for evaluating placement decisions. In the case of exemption, the treatments are exemption and non-exemption; the trait variable is a measure of prior knowledge; and a measure of educational benefit is proposed as the criterion for evaluating performance under the two treatments.

Table 2: Distinction between the character and use of placement and exemption tests.

	PLACEMENT TEST	EXEMPTION TEST
Purpose emphasized	To establish point of entry in a course sequence	To waive a course requirement
Nature of decision	Tentative, student may participate	Usually final, faculty determines
Credit awarded	Usually not	Generally yes
Cutting score	Liberal (more students are advanced)	Conservative (fewer students are advanced)
Norms required	Local	Local, systemwide, or class of institution
Content of test	Closely tuned to course, usually diagnostic	More general measure of comparable competency
Structure of test	Shorter, more numerous subscales, especially in a segmented sequence	Longer, more reliable test; usually little need for subscales
Need for security	Less	More

Willingham, Warren W., College Placement and Exemption. New York: College Entrance Examination Board, 1974, p. 97.

Since, in both placement and exemption, an achievement test serves as a trait assessment, some assurance of content validity is needed before the test is administered to the students for the purpose of obtaining local norm data. This content validity is frequently established by having subject matter experts review the content of the test. Placement decisions are more closely tied to instruction at the local institution than are exemption decisions. Students being placed into a sequence will be expected to perform as if they had taken the prerequisite local course(s). This is not generally true in the exemption model. In this case, the student is exempted on the basis of previous knowledge and this exemption is treated in much the same way as if the student were transferring credit from another institution. Thus, content validation needs to be specific to a particular course at the local institution in the case of placement, but in the case of exemption a more general content validation is recommended. That is, for exemption the question to be answered is more like "Should we accept transfer credit/exemption based on the content of the test?" and in the case of placement "Does this content specifically match the local course?". In either the placement or the exemption model, if there are several instructors teaching the course in question, some degree of agreement in judgements and in course content is assumed.

> RECOMMENDATION 2: To assure content validity for a placement test, if several instructors teach the course under consideration, each should independently review the test item by item to determine whether the content and skills covered in the item are covered in his/her version of the course.

After content validity has been established by having subject matter experts examine the test, it may be checked empirically by

administering the test to students completing the relevant course and computing a correlation between the test score and a measure of course performance.

MODELS FOR PLACEMENT AND EXEMPTION TESTING

Three placement models are recognized; placement within a sequence, pacing, and remediation. Placement within a sequence is commonly used in foreign languages, a calculus sequence in mathematics, and any other subject in which one course is a pre-requisite for the other. Pacing is distinguished by keeping the amount of content material constant, but varying the length of time spent on the material. Thus, an advanced course may cover the same content as another course but in half the time or, at the other end of the scale, in twice the length of time. The third model for placing students into alternate treatments is for the purpose of remediation. This situation occurs in some sciences (e.g., Chemistry) which require certain mathematical skills, or a calculus course which requires a good understanding of functions (usually covered in a pre-calculus course), or in an English Composition course which requires an ability to use proper grammar, punctuation, and spelling. In each of these, the choice is to place students into a long sequence (including the remedial course) or a short (standard) sequence.

In the case of exemption we restrict our attention to a single model. A single required course is under consideration (as opposed to a sequence of courses) and the decision to be made is to exempt the student from the course (with or without credit) or to deny exemption.

Thus, there is no logical second course for the students to take. Exemption is to be made on this basis of prior knowledge.

APPLICATION OF TTI TO PLACEMENT AND EXEMPTION MODELS

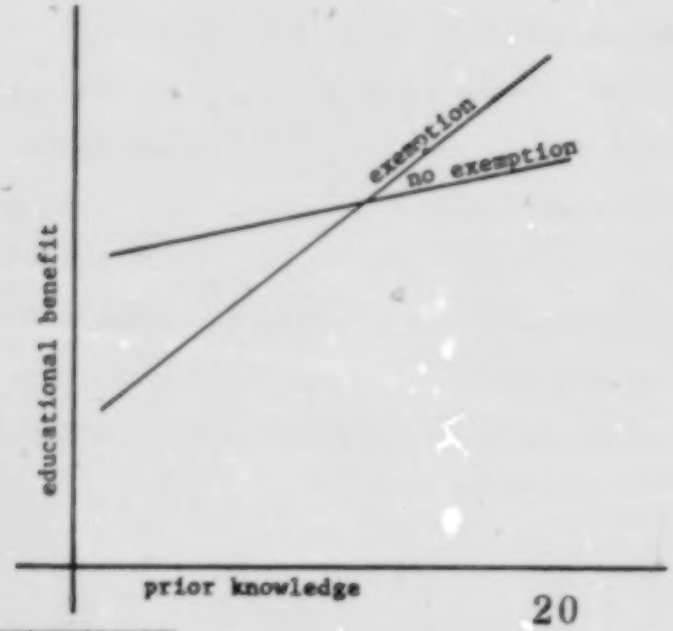
In the three placement models designated earlier, the choice is between placing the student into a long or short sequence. TTI in the case of placement is illustrated in Figure 4. If the test is valid for the placement situation under consideration, then students with low scores on the test should be placed into the long sequence and students earning high scores should be placed into the short sequence. Thus, a choice for the cutoff score for differentiating between placement into the long or short sequence is that placement test score corresponding to the intersection point of the two lines. It is assumed that the slopes of both lines are positive.

Figure 4: TTI for Placement Models

In Figure 4 the criterion variable is given as course achievement. Students who would be placed into the short sequence save time, possibly tuition, or can take a course of their choosing in the time saved from the required course. These cost-benefit factors can be included to obtain a figure similar to that illustrated for the exemption model in Figure 5. The regression lines representing either course is adjusted up or down depending upon whether the net utility is positive or negative* to obtain educational benefit (or utility) as the criterion instead of second course achievement.

In the TTI model for exemption illustrated in Figure 5, the treatments are exemption and non-exemption from the course under consideration, the trait variable is a measure of prior knowledge which is the exemption (achievement) test, and the criterion is a measure of educational benefit. The score on the trait variable corresponding to the intersection of the two lines serves as a logical cutoff score for awarding exemption.

Figure 5: TTI for exemption models.



*Assuming that the net utility is the same at each trait level.

METHODS FOR DETERMINING CUTOFF SCORES

The usual methods for determining whether a test to be used for placement or exemption covers appropriate content is to compare the content of the test with that of the prerequisite course and/or correlate scores on the test with achievement at the end of the prerequisite course. If content validity is sufficient, the data obtained by either of these methods can be used in a variety of ways to choose the cutoff score. When content validity is evaluated by having experts in the subject matter area compare the content of the test with that of the course, we will illustrate how the cutoff score can be determined using that content match alone or in conjunction with national norm data. When the correlational approach is used to evaluate content validity, we will show how various regression methods or discriminant analysis can be used to determine an appropriate cutoff.

The seriousness of the potential decision errors may also affect the choice of the cutoff score for placing students into the long or the short sequence. The relative seriousness of the errors will depend in part upon whether a score below the cutoff results in <u>requiring</u> or in <u>recommending</u> the longer sequence.

The methods for choosing cutoff scores discussed in this paper are classified according to the type of data used in each. Each method will be examined and illustrated with empirical data when data are available and fit the situation. In the final section of the paper, a method

for evaluating these cutoff scores in terms of TTI is proposed. Recommendations based on our experiences are scattered throught this part of the paper.

NO EMPIRICAL DATA

The first method assumes that the cutoff score must be chosen based on the content of the test in the absence of national or local norm data. Usually the decision to place students is made before the placement test has been chosen or constructed locally and before the test has been validated. Thus, it is not uncommon to be required to set a cutoff score in the absence of norm data, perhaps temporarily until a validation can be designed and the cutoff score evaluated. Meskauskas (1976) concludes his review of such methods with the statement "Of the approches reviewed in this paper, only the Nedelsky method has received extensive practical use. The other methods need to be validated to provide users with data on which to make choices." (p. 155) Thus, we will restrict our attention to the Nedelsky method.

Nedelsky (1954) developed a method for establishing an absolute standard for the minimum pass level (MPL) for objective type tests. The method has been applied primarily to choosing a cutoff score for passing a criterion-referenced test, but also seems useful for choosing cutoff scores for placement or exemption. To use this method, each instructor of the course under consideration (prerequisite course or the course for possible exemption) carefully considers each item of the test and decides which distractors the lowest passing student should be able to identify as incorrect. For each item, the item MPL is the

reciprocal of the number of remaining distractors. Thus, if the item has five options and the lowest passing student should be able to recognize two of the options as being incorrect, the MPL for that item is 1/3. The MPL for the entire test is the sum of the item MPLs. Since each instructor determines an MPL, these individual MPLs for the test can be averaged to obtain the departmental MPL. The final MPL obtained depends upon the differences of opinion of the instructors who evaluate the test regarding what is adequate performance and upon the number of instructors involved in the standard setting. Nedelsky proposed a method for incorporating the standard deviation of the instructor ratings into the final MPL, but it doesn't seem to have worked well in practice and most uses of the method involve only the average instructor MPLs. Other researchers have found that MPLs should be based on opinions of more than one faculty member. From the validation of the method by other researchers, the method seems to be useful, but more work remains to be done in this area.

> RECOMMENDATION 3: Cutoff scores determined using only test content (such as Nedelsky's MPL) should be based on the opinion of several subject matter experts.

EXTERNAL NORM DATA

Suppose that, in addition to copies of the test itself, norm data is available for a relevant norm group. We present two ways to interpret the external norm data.

Choosing Cutoff Score Based on National Norms. If the percentile ranks of the scores are provided, an equal percentile method can be used to obtain a cutoff score. This method is applied by choosing a

criterion measure which is available for the relevant group at the local institution. This criterion may be a measure such as the score on a comprehensive final examination for the prerequisite course or a final course grade. For example, frequently a student with a grade of C or better is judged to be adequately prepared for the next course in the sequence, in which case the percent of grades below C recommended or typically awarded could serve as the percentile cutoff for placing students into the next course. The test score corresponding to that percentile is chosen as the placement test cutoff score. This, if the recommended grade distribution for the institution designates that, on the average, 11% of the assigned grades be D or F, then the test score corresponding to the 11th percentile would be chosen as the cutoff for placement into the next course. In the case of exemption, often the lowest B is chosen as the standard for awarding exemption based on the test score. The procedure is applied in the same manner as for placement. This method incorporates a departmental or institutional choice of what grade level constitutes adequate preparation for the next course and institutional values in assigning grades.

Sometimes a test publisher does not provide national norm percentiles, but does indicate the performance of a typical A student, typical B student, and so on. Making some assumption on the distributions of grades, the average score earned by C and D students provides an estimate of the score earned by the lowest C student which could be used as the cutoff score. A better method might be to use a weighted average of scores earned by C and D students; the weights being the proportion of students in the particular institution recommended to receive the grades of C and D.

Minimum Recommended Cutoff Scores Based on Chance Performance. Scores earned by students who are randomly guessing have a distribution with an expected chance score and a standard deviation. It is common to consider the expected chance score and expected standard deviation of such scores when studying a multiple-choice test. Generally, scores which are more than two standard deviations from the expected score (mean) are considered to be so far from the mean that it is likely that they are from a different distribution. Thus, those scores which are more than two standard deviations above the expected chance score can be considered to arise from other than random guessing. We suggest that the score corresponding to two standard deviations above chance is the minimum acceptable cutoff score for either placement or exemption.

The expected value and standard deviation of chance scores is calculated as follows. Let p represent the probability of getting the item correct by guessing; this is $\frac{1}{\text{number of options}} = \frac{1}{4}$ for one Modern Language Association (MLA) Cooperative Foreign Language Achievement Test series we were considering. Let n represent the number of items on the test. Then the expected (or mean) chance score is the product np and the standard deviation is $\sqrt{np(1-p)}$. Thus in the case of the listening subtest for level M, there are four choices for each item and 40 items in the test, the expected score earned by guessing is $(40)(\frac{1}{4}) = 10$, the standard deviation of chance scores is $\sqrt{(40)(\frac{1}{4})(\frac{3}{4})} = 2.7$. So using the two standard deviation rule, then the minimum cutoff score for placing students beyond second year of the language (or, in many cases, exempting them from the university language requirement) should be greater than 15.4. Two levels of the MLA tests are available, Level L for students

completing one year of the foreign language and Level M for students completing two years of the language. In addition to two levels of each language, there are two forms, Form A and Form B of each level. The number of items in each subtest is the same for the three languages of interest to us, French, German, and Spanish, and for each form of the test. However, the listening subtest for Level M contained 40 items whereas the listening subtest for Level L contained 45 items. In Table 3 we present the expected value and standard deviation of the

TABLE 3: Minimum cutoff scores based on the mean and standard deviation of chance score.

Level and Subtest	Number of Items	Expected Chance Score and Standard Deviation*	Expected Chance Plus Two Standard Deviation	Estimated PR of Previous Column**	
				Form A	Form B
Level L					
Listening	45	11.25	17.05	37	37
		2.9		36	30
				36	31
Reading	50	12.5	18.7	20	15
		3.1		33	23
				29	29
Level M					
Listening	40	10.0	15.4	65	60
		2.7		28	63
				38	47
Reading	50	12.5	18.7	45	33
		3.1		57	54
				57	61

*All items have four options

**Entries are French, German, Spanish

chance score, the minimum cutoff raw score based on the above assumption, and the estimated percentile rank for that score based on the national norm data supplied by the test publisher. For example, in the case of the French listening subtest for Level M, we can only be relatively sure that 35% (100-65%) of the national sample were not randomly guessing. For Level M of the three languages under consideration the minimum recommended cutoff scores correspond to the 28th to the 65th percentiles and for seven of the twelve tests the minimum cutoff score is above the median. Thus, in the case of Level M, the results should be cause for some caution in using the test for students in the midst of or completing their second year of the language in college. For either placement or exemption, the cutoff score should be at least sufficiently high to provide assurance that a student achieving at the cutoff is exhibiting above-chance performance.

> RECOMMENDATION 4: A cutoff score for placement or exemption should be at least two standard deviations above the expected chance score.

General Strategies for Placement and Exemption. Placement is an instructional strategy and thus needs to be closely related to the course. The emphasis in placement is on placing students into a course rather than granting exemption from a course, although the latter often occurs as a consequence. Since textbooks and course emphases differ among institutions, a test that is valid for placement at one institution is not necessarily valid at another (since different aspects of the subject may be emphasized in the courses in the sequence). Thus, we do not recommend placing students into one or other course on the basis of national norm data alone.

RECOMMENDATION 5: Placement decisions should be based on local data and not solely on national norm data.

Exemption decisions, on the other hand, are made to grant the student credit for knowledge gained elsewhere, in a manner similar to credit granted for knowledge gained at another institution. Thus, the exemption decision may be based solely on national norm data and content validation by subject matter experts at the institution.

RECOMMENDATION 6: Cutoff scores for exemption may be based on national norm data and local content validation.

LOCAL NORM AND CRITERION DATA

Local norm data is obtained by administering the test to a selected group of students at the local institution. Appropriate criteria data (e.g., score on a comprehensive final examination, course grade) for these students should also be gathered when possible. In this section we discuss several methods for using local data to determine a cutoff score.

Local Norm Data Without Criterion Data. When local norm data are the only local data available (criterion data are not supplied or are inadequate) the method outlined in the case of national data only can be applied. The data presented in Table 4 are local and national data on CLEP Educational Psychology examination recently approved as the basis for awarding credit by examination and exemption from the course requirement at the University of Iowa. The faculty felt that with the problems of grade inflation, the cutoff score should be based on Educational Policy Committee recommendations that, on the average, 36% of students in such courses should be receiving A's and B's. This value of 36%

Table 4: Scores of University of Iowa students on the CLEP Educational Psychology subject examination.

Scaled Score	Frequency	Percentile Rank (Local)	Percentile Rank (National)
74	1	99	99
71	1	98	99
69	1	96	97
66	1	95	94
64	1	93	91
63	1	92	89
62	4	88	87
61	4	82	85
60	3	76	83
59	1	73	80
58	2	71	77
56	2	68	72
55	3	64	69
54	4	60	65
52	2	55	58
51	5	49	55
50	2	44	51
49	2	41	47
48	5	36	43
47	1	31	40
46	2	29	36
45	5	26	32
44	4	17	29
43	3	11	26
41	2	8	21
40	2	5	18
39	1	2	16
35	1	1	7

MEAN 51.8

STANDARD DEVIATION 8.6

Based on 66 students completing 7P:75 (Educational Psychology and Measurement) at the University of Iowa in fall, 1975.

led them to choose a cutoff score, based on the local norm data, of a scaled score of 55.

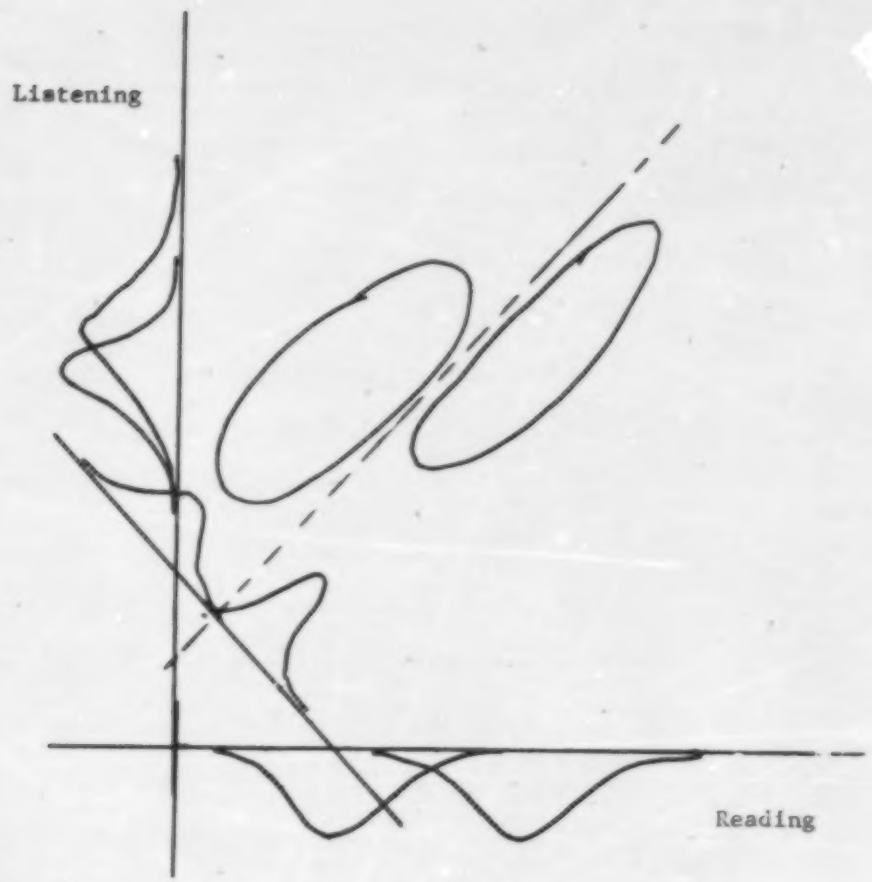
Discriminant Analysis Procedure. When the test under consideration is to be used to determine whether students are to be placed into first or second semester of a sequence (e.g., language), the test could be administered simultaneously to all students (or a random sample) completing first and second semester or at the beginning of a second and third semester. If the test is valid for placement, at the very least, it should discriminate between students completing one semester and those completing two semesters of the language. (In addition, the latter should earn higher scores on the average.)

Consider the situation where test data are available for students who are completing the first (prerequisite) course and for different students completing the second course. That is, the test was administered simultaneously to students completing both courses. In this case a discriminant analysis type procedure can be used to choose as a cutoff score which will minimize the number of misclassifications. This often amounts to choosing the average of the two means as a cutoff score. If information on the seriousness of the errors in classifications is available, this can be incorporated into the procedure to adjust the resulting cutoff score. This information may take into account that one type of error is easier to correct than the other; or the fact that the decision to be made is that the first course is recommended to the student (rather than being required).

When the test contains subscales such as the Modern Language Association (MLA) tests which contain the four parts of reading, listening, speaking, and writing, the cutoff score can be based simply on

the total score, a multivariate cutoff can be determined, or cutoff scores can be determined on each subscale; a discriminant analysis procedure used in each case. In general, the multivariate cutoff will yield the highest rate of correct classifications, but may not hold up as well under cross-validation. Figure 6 illustrates one possible outcome using two of the MLA subscales, reading and listening.

Figure 6: Discriminating between two groups using a test with two subscales.



Notice that even though the two subscales are correlated within each course, using the two scales will result in a clearer separation of the two distributions than either scale used alone.

To illustrate, we administered the MLA Spanish, Level L test to students who were completing the first, second, and third semesters of Spanish. Results of applying discriminant analysis to pairs of groups are presented in Table 5.

Table 5: Percent of students correctly classified using discriminant analysis of MLA Spanish Level L.

Semesters of Spanish Completed	Reading Subscore	Multivariate Read-Listen	Total Score	N
1 and 2	51.27	60.53	51.64	304
2 and 3	71.00	75.00	81.98	172

The first row of the table contains the results for discriminating between students who have had one and those who have had two semesters of Spanish. Using the discriminant analysis procedure (which minimizes the number of misclassifications) only about 50 to 60 percent of the students in the validation sample are correctly classified, and the multivariate procedure yields the highest percent of correct classifications. We expect that when this cutoff is applied to a new sample the percent correctly classified will be lower. From these results the test does not seem valid for discriminating between students completing one or two semesters of Spanish. The test does seem to distinguish well between students completing two or three semesters of Spanish with 70

to 80 percent of the validation sample correctly classified. An additional conclusion based on the results in Table 5 is that the listening test is valuable in discriminating between two or three semesters of Spanish. Using only the reading subscore results in the lowest percent correct classifications.

An inspection of the means and standard deviations of test scores for the various groups (Table 6) supports this conclusion.

Table 6: Means and standard deviations obtained for local administration of MLA Spanish Level L.

	Listen	Read	Total
Chance	11.25	12.5	23.75
	2.9	3.1	4.2
First Semester	16.7	19.1	35.9
	6.0	6.6	11.6
	(169)	(169)	(160)
Second Semester	16.7	22.2	38.9
	6.5	6.8	12.1
	(166)	(146)	(145)
Third Semester	24.9	28.3	53.2
	7.6	9.8	16.2
	(35)	(35)	(28)

Test was administered to students completing one, two or three semesters of Spanish.

Relating these results to those in the previous section, we see that the mean listening score for both first and second semester students is less than two standard deviations above the expected chance score. However, the reading subscore is above the range of chance scores we have established. Thus, if we can assume that these results are representative, the MLA Spanish Test Level L is not valid for determining placement into first or second semester Spanish at Iowa.

Equal Percentile Method. Suppose that test data is obtained only for students completing the first (prerequisite) course. The equal percentile method outlined in the discussion of national norm data can be applied to the local norm and criterion data. If grades are used as the criterion and the lowest C is chosen as the standard, this method is applied by counting the number of students in the sample who earned a course grade below C. Order the test scores from lowest to highest and count that number of scores starting at the bottom of the distribution, choosing as a cutoff the lowest score remaining. This method is sometimes called the equivalent score method since the score obtained is equivalent in one sense to that earned by the lowest C student.

In placement, frequently the lowest C is chosen as the standard for placing students into the next course. Other standards have been proposed and applied. For example, Aleamoni (1973) recommends choosing a cutoff score for placement such that practically no students who earned a D or F in the course exceeded it. This standard typically yields a higher cutoff score than using the lowest C level.

In the case of exemption, the methods described in the previous paragraph can be used (adjusting the standard if necessary). Applying the equipercentile method to the Educational Psychology data (exemption) presented earlier (Table 4), 43 students received an A or B in the course so that a scaled score cutoff of 48 is obtained. Note that this result differs substantially from that obtained earlier using local test results without the criterion data. In this case it occurs because faculty awarded more A's and B's than recommended.

Regression Method. A second approach to choosing the cutoff score in this situation is that of regression. The advantage of regression over the equal percentile approach described above is that all scores (not just those earned by students with grades below C or those earned by students with C or D grades) are used to determine a cutoff score. Two linear regression methods and a compromise will be discussed.

Regression is usually used to predict one score from another. In the placement and exemption context, the test and course criterion are concurrent measures, so we can predict the course criterion (such as final course grade) from the placement test score, or predict the placement test score from the course criterion. These two approaches do not generally yield the same result. A typical outcome is illustrated in Figure 7. Suppose we wished to choose as a cutoff that score corresponding to the lowest C, or a grade of 1.5. The intersection point of the two lines is the average grade and the average test score. In Figure 7, the average grade is assumed to be C (2.0) and the average test score 50. The cutoff obtained using the test regression line is lower than that obtained using the grade regression line. A line bisecting the angle between the two regression lines is suggested as a compromise to establish the cutoff score. Most methods for choosing cutoff scores are based on one of these three regression lines (Willingham, 1974, p. 170).

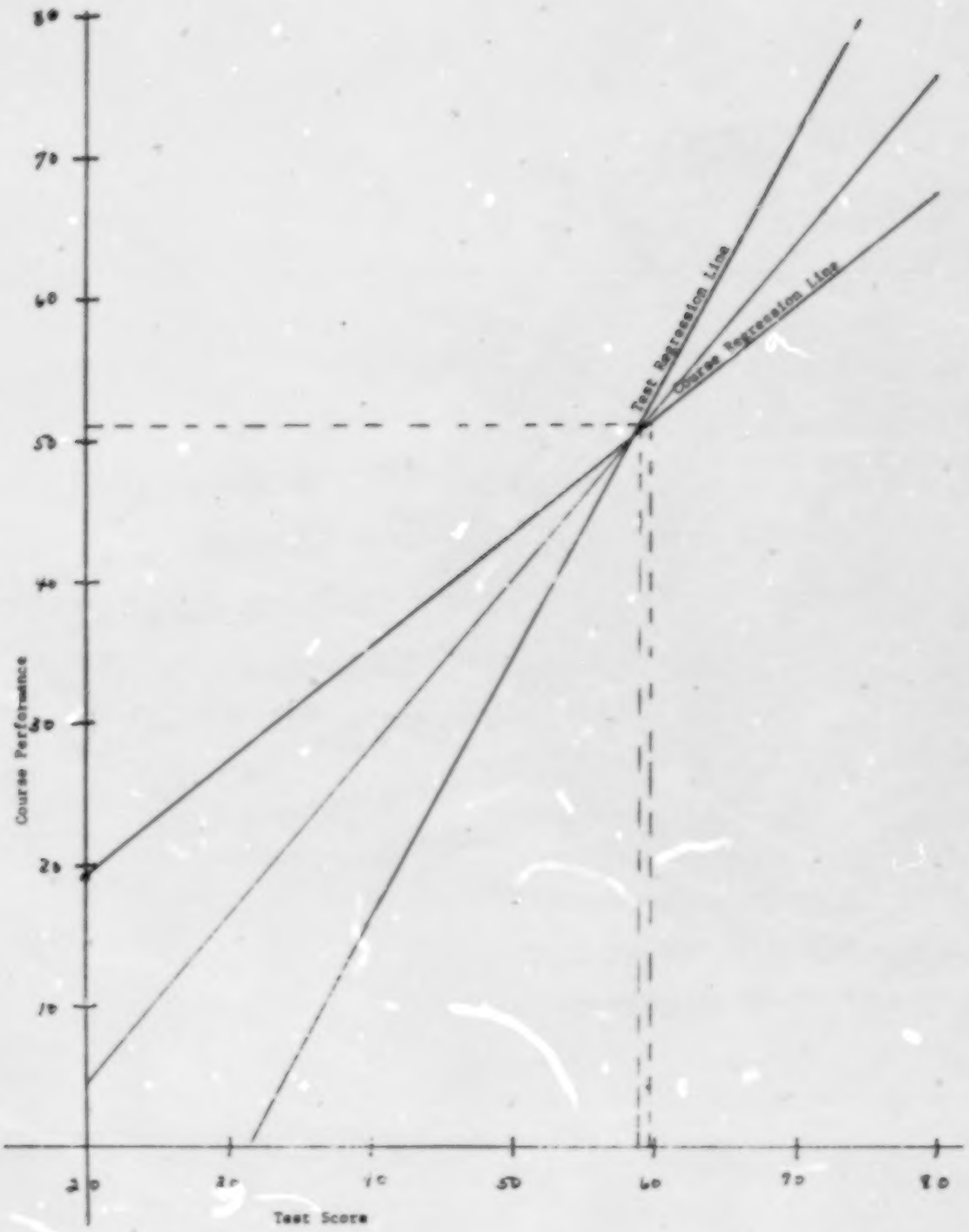
Figure 7: Regression lines for choosing cutoff scores.

In Figure 8, these methods are applied to the local data on the CLEP Introductory Sociology examination administered to students completing a one-semester sociology course. However, course grade is replaced by course performance as the criterion. The course performance variable represents the total score obtained on the instructor-constructed exams taken during the semester scaled within each section to have mean 50 and standard deviation 10. Each section had a different textbook and took different exams, so the scaling within section was performed to enable comparisons across sections (the smallest section in the study had approximately 100 students). If the decision-maker feels that, say, 46% of the students should be receiving A's and B's, and a course standard of 51 attains that corresponding percentile, a cutoff score near 59 or 60 would result. Similarly, this method can be used to choose a cutoff score for placement decisions.

RELATIVE SERIOUSNESS OF CLASSIFICATION ERRORS.

Suppose that, in addition to local test and criterion data, it is possible to identify the relative seriousness of the errors made by misclassifying students. This information can be incorporated into the discriminant analysis procedure as described above, or into the regression method by choosing a higher or lower grade level for establishing a cutoff score. We describe a third method incorporating these as losses into the procedure for choosing the cutoff score. As in the case of the regression procedures discussed above, a decision must be made for determining, on the basis of the criterion, who should be exempted and who should not.

Figure 8: Applying the methods for choosing a cutoff score for CLEP Introductory Sociology based on course performance.



In the validation study, when criterion data are available, we form a table (Table 7) and calculate the proportion of the sample in each cell of the table. For example, using the Educational Psychology data introduced earlier, if we choose total score earned on classroom tests as our criterion, and the mean

Table 7: Decision theory model applied to the validation sample.

		Criterion Data	
		Above Standard	Below Standard
Test Data	Above Cutoff	Correct Decision	Error
	Below Cutoff	Error	Correct Decision

as the standard, we can form a table such as Table 7 for each possible cutoff score on the test. These values are presented in Table 8 for selected cutoff scores. Using a cutoff score of 50 results in granting exemption to 35% (.349)

Table 8: Table of estimated* proportions for cells of Table 7 based on Educational Psychology students at the University of Iowa.

Cut-off Score	Successes That Will Be Selected	Failures That Will Be Rejected	Failures That Will Be Selected	Successes That Will Be Rejected
70	.071	1.000	.000	.929
65	.107	1.000	.000	.893
60	.464	.968	.032	.536
55	.607	.839	.161	.393
50	.756	.651	.349	.244
45	.902	.331	.669	.098
40	1.000	.065	.935	0.000
35	1.000	0.000	1.000	0.000

Success is defined as above the mean on classroom tests.
Using this definition of success, 48% of the sample is classified as successful.

* Data was smoothed to emphasize the positive relationship between classroom test and CLEP test scores. (See Perrin and Whitney, 1976.)

of the students who are below the standard while denying credit to 24% (.244) of the students with criterion scores above the standard. As the cutoff score is raised, fewer failures will be selected but a greater proportion of successes will be rejected. Thus, it is useful to consider the relative seriousness of these errors. There are two possible errors, granting exemption to an undeserving student or denying exemption to a deserving student. Let α and β represent the loss associated with each of the two errors. Without loss of generality, we can assume that these are scaled such that $\alpha + \beta = 1$. α and β can be interpreted as losses associated with making the corresponding errors.

Suppose that the faculty members who are responsible for making the decision of a cutoff score for awarding exemption feel that it is three times as serious an error to award exemption to an undeserving student as to deny exemption to a deserving one. In this case α and β would be scaled to .25 and .75 so that $\alpha + \beta = 1$. The expected loss associated with each possible cutoff score can then be calculated by weighting α and β by the observed relative frequency with which each occurs in the sample. For example, the expected loss associated with choosing 55 as a cutoff score (in the 3 to 1 case), is (.161) (.75) + (.393) (.25) = .219. Whichever potential cutoff score yields the minimum expected loss is chosen as the cutoff. Values obtained for selected losses and selected cutoff scores based on results in Table 8 are presented in Table 9.

Suppose it is reasonable to assume that "above the mean on classroom tests" is an adequate definition of success. Then, in retrospect, by choosing 55 as the cutoff score on the CLEP test for awarding credit, the Educational Psychology faculty's action indicates that the relative seriousness of the errors is in the neighborhood of one to one -- that is, it is equally as serious an error to award exemption to an undeserving student as to deny exemption to a deserving one.

Table 9: Expected losses resulting from using selected cutoff scores for awarding exemption to Educational Psychology students.

	3 to 1	2 to 1	1.5 to 1	1 to 1	1 to 1.5	1 to 2	1 to 3
70	.232	.307	.372	.465	.557	.622	.697
65	.223	.295	.357	.447	.536	.598	.670
60	.158	.198	.233	.284	.335	.370	.410
55	.219	.238	.254	.277	.300	.316	.335
50	.323	.315	.307	.297	.286	.278	.270
45	.527	.480	.440	.384	.327	.287	.241
40	.701	.626	.561	.468	.374	.309	.234
35	.750	.670	.600	.500	.400	.330	.250
Minimum Average Loss	.158	.198	.233	.277	.286	.278	.234
Cutoff Score	60	60	60	55	50	50	40

The method as described here for exemption can be used for placement as well. Generally, in the case of placement it is not as serious an error to place a student too high as it is to grant exemption to an undeserving student, since the former are usually easier to correct. If the situation is one of deciding whether to recommend to a student that he/she take a mathematics course before the chemistry course the ratio may even be reversed. That is, it may be better to make such a recommendation when it is not necessary than to fail to make it when it is necessary.

This discussion has been limited to using the same loss for all errors in a particular direction. Livingston (1975) suggests using a linear loss function. Any reasonable loss function can be used.

OTHER RELATED MEASURES

To Establish the Decision Rule. Occasionally, other measures related to success in the criterion course are available in addition to the local test data. These may include such measures as number of semesters of high school

courses taken in the subject matter area, reason for taking the course, college entrance test scores, or college grade point average in the subject courses. These measures can be used as additional discriminators in the discriminant function or as additional predictors in the regression equation. The cutoff score will then depend upon relevant discriminators or predictors. Sometimes it may be necessary to restrict attention to a particular subgroup and base the cutoff score on results obtained for part of the total group.

To Explain a Result. Another reason for gathering additional information which is not explicitly part of the research design is to help explain ambiguous results. For example, we administered the CLEP Introductory Sociology subject examination to students who were beginning the introductory sociology course on campus, with the result that 40% achieved a score above the current cutoff (median of the national sample) for awarding credit (and exemption) at the University of Iowa. At the time they took the test, the students indicated the number of semesters of sociology they had taken in high school. We found that while previous coursework and ACT Social Studies scores were relatively independent, both helped explain the high CLEP scores earned by students. That is, on the average, students who had sociology in high school performed better on the test than students who did not have previous coursework in sociology. We also found a moderate correlation between CLEP Introductory Sociology test score and ACT Social Studies subscore ($r = .59$).

As a second example illustrating the value of gathering related information, we use data collected on the try-out of items for the CLEP German subject examination. The test was administered to students completing one or two years of German. We administered 60 items in the area of reading and 54 items in

the area of listening, all were four-option multiple-choice items. For internal curriculum evaluation, we compared scores earned by students who had chosen a program which emphasized listening with scores earned by students who had chosen a program with reading emphasis. Results obtained for students completing the second year are discussed here. Students in both programs took the reading test. A test of hypothesis for equal means for the two programs yielded results which indicate that the means are different. It appears that the students in the listening program can read German better (mean = 34.5) than the students who are in the program which emphasizes reading (mean = 28.9).

During the administration of the CLEP German Pre-Test the students were requested to answer the following question.

> Which of the following reasons most nearly explains why you are taking German? (Mark only one.)
> (1) to satisfy a University requirement.
> (2) to satisfy a departmental requirement.
> (3) to get elective credit.
> (4) I plan to major in German.
> (5) I plan to minor in German.
> (6) because I like it and wish to become more fluent.
> (7) other (please specify on back of your answer sheet).

To aid in interpreting the results these reasons were classified into three categories: requirement (1 or 2), elective (3, 6, 7), major-minor (4 or 5). We anticipated that students who were taking the course for a requirement would tend to perform less well on the test and in the course than students who were taking the course as an elective. Furthermore, we expected German majors and minors to be highest in ability for German and most highly motivated both in the course and on the test.

We recognize that second year students are not always sure whether they will major or minor in an area, or if they are just fulfilling a liberal arts requirement, but it seemed that perhaps students who were taking the course as a requirement would tend to choose the reading program as opposed to listening.

Students would probably perceive the course with listening emphasis to be more difficult because of the necessity of mastering an additional skill. Similarly, those taking the course as an elective or with a plan to major or minor in the area might tend to choose the program with listening emphasis. Thus, the scores were broken down by reason for taking German and are presented in Table 10.

Table 10: CLEP German Pre-Test results for second year students broken down by program choice and reason for taking German.

Course	Reason for Taking Course	Listening Subscore	Reading Subscore	Total Score	Course Grade
Regular Listening	Total Group	25.3 10.2 (21)	35.4 9.5 (21)	61.0 18.4 (20)	3.0 1.3 (22)
	Requirement	15.7 8.6 (3)	28.8 6.2 (4)	41.7 9.8 (3)	2.2 .5 (4)
	Elective	26.7 10.1 (15)	35.8 10.0 (14)	63.0 18.5 (14)	3.1 1.4 (15)
	Major-Minor	28.0 9.2 (3)	42.7 4.7 (3)	70.7 13.5 (3)	3.3 .6 (3)
Regular Reading	Total Group		28.9 13.1 (21)		3.0 1.1 (21)
	Requirement		27.5 11.0 (16)		2.9 1.1 (16)
	Elective		33.4 19.3 (5)		3.4 .9 (5)

Entries contain mean, standard deviation, and the number of cases on which the calculations are based.

The sample sizes for the two courses were small (21 in each), and after breaking down results by reason for taking the course, sample sizes are as low as 4. Thus, generalizations from the results presented in Table 11 are not recommended. However, for the purpose of illustration, we observe that although the means for the two courses are different when studied as a whole (28.9 and 35.4), that difference does not appear when attention is restricted to either the students who are taking the course as a requirement (27.5 and 28.8) or those who are electing to take the course (33.4 and 35.8). The difference appears to reflect primarily the difference in the proportion of each in the two programs.

For Future Counselling. A third reason for gathering related information is for the use in future counselling of students. In the study of students enrolled in the beginning Sociology course, discussed earlier, a correlation of .59 (N = 646) between American College Test (ACT) Social Studies score and the CLEP Introductory Sociology test administered prior to the course was noted. This indicates that the ACT subtest is somewhat useful in predicting success (whether or not students receive credit) on the CLEP test.

Table 11 was prepared to help advise students on the likelihood of receiving CLEP credit in Introductory Sociology based on ACT Social Studies

Table 11: Expectancy of Success on the CLEP Introductory Sociology Test Predicted by the ACT Social Studies Score.

ACT Social Studies Score	Estimated Probability of Earning Credit
3-7	.07
8-17	.09
18-20	.19
21-23	.30
24-25	.39
26	.60
27	.67
28	.75
29-30	.77
31-33	.82
34	.99

score. ACT scores are grouped to provide better estimates of the probability of earning credit and to emphasize the positive relationship between ACT and CLEP scores. The entries are based on the median of the national sample as a cutoff score for awarding credit.

Thus we have found additional information to be helpful, not necessarily in establishing the decision rule, but in explaining a result or for advising students regarding their chances for success. The information can be used in making the decision when, after the data has been gathered, it is learned that one subgroup is different from the others.

> RECOMMENDATION 7: We recommend gathering additional information from students whenever possible for the purpose of explaining ambigous results. This information may include: number of semesters of previous coursework in the area, reason for taking the course, whether or not they took another specific course and grade in that course, what grade they feel they deserve in the course in which the test is being administered.

RELATED ISSUES

Motivation. We conclude this discussion of methods for choosing cutoff scores with a discussion of a problem inherent in all local test administration; namely, student motivation. This issue seems to be largely ignored in testing literature. The recent emphasis and concern regarding experimentation with human subjects demands that the motivation be such that it encourages students to participate and that they take the task seriously. Motivation is critical if the results are to be representative. The course instructors need to be involved so that adequate motivation is provided.

Some suggestions for providing motivation for a pretest follow. If the course in which the student is enrolled is not to be used in the validation process (using previous semester grades for validation), it would be possible to offer the students exemption from the course for a score above a previously determined standard. In some situations, posting scores with the identification numbers of students who achieved above a given point (such as the median of the national sample) would provide the necessary motivation. During the first week of a beginning sociology course, we administered the CLEP Introductory Sociology test to validate it for credit by examination to learn how much the students know at the beginning of the course. As a motivation, we sought and received approval to offer the students one semester hour of ungraded credit in Sociology for a sufficiently high score on the test. Since this examination is currently being used for credit by examination at our institution, the cutoff for the one hour of ungraded credit was the current cutoff score for awarding credit by examination.

Motivation for a posttest usually takes the form of being part of the student's final course grade. Since almost any test used for placement or exemption will cover some material that the instructor did not cover in the course and the course will contain some topics not covered on the test, instructors are justifiably concerned about using the test as THE final course examination. The best procedure seems to be to use the test as part of the final examination with additional items written by the instructor. The two parts are weighted according to content coverage in the course; the score on items from material not covered in class used only to raise the students grade. The instructor may tell the students that their score on this test can only help their grades. However, we do not judge this practice to provide adequate motivation for the entire group of students. Another form of the Introductory

Sociology test was administered to students at the end of the course. One instructor used the score on the posttest as part of the grade and another instructor used it only to improve the grade. Although the mean scores on the pretest were comparable, the mean scores for the sections at the end of the course differed by one-half standard deviation. This may have been due simply to differences in the motivation described above.

> RECOMMENDATION 8: The problem of student motivation needs serious attention in planning local validation administration of a test to be used for placement or exemption.

Sample Size. When sample sizes are small, correlations are greatly influenced by sampling error. Sample sizes of 100 per group are generally recommended if the resulting decisions are to be based on sample correlation coefficients. This restriction limits the generalizability of the data when groups are subdivided on further related variables.

In the past, obtaining adequate sample sizes did not pose a serious problem in educational research. Recent guidelines regarding experimentation with human subjects sometimes apply to validation administrations of tests to be used for placement and proficiency. This can create difficulty in obtaining a sufficiently large random sample for the research. This problem is often alleviated when the test is examined by the course instructors and then used as a part of the course. This seems a preferable approach in most situations.

Grades. We have not found grades to be an adequate criterion for evaluating course performance when validating tests to be used for placement or proficiency. Among the problems we have encountered are: a five-point scale is too restricted, grades depend very much upon the individual instructor's philosophy of evaluating course performance, grades sometimes depend upon factors extraneous to achievement, and the test being validated may be used in computing

the grade. Thus, we prefer to use another measure of course performance. When available, score on a comprehensive final examination or total number of points earned in the course seem to provide a better means for evaluating course performance than a letter grade.

> RECOMMENDATION 9: A measure of course performance other than final course grade should be used as a criterion for test validation and establishment of cutoff scores.

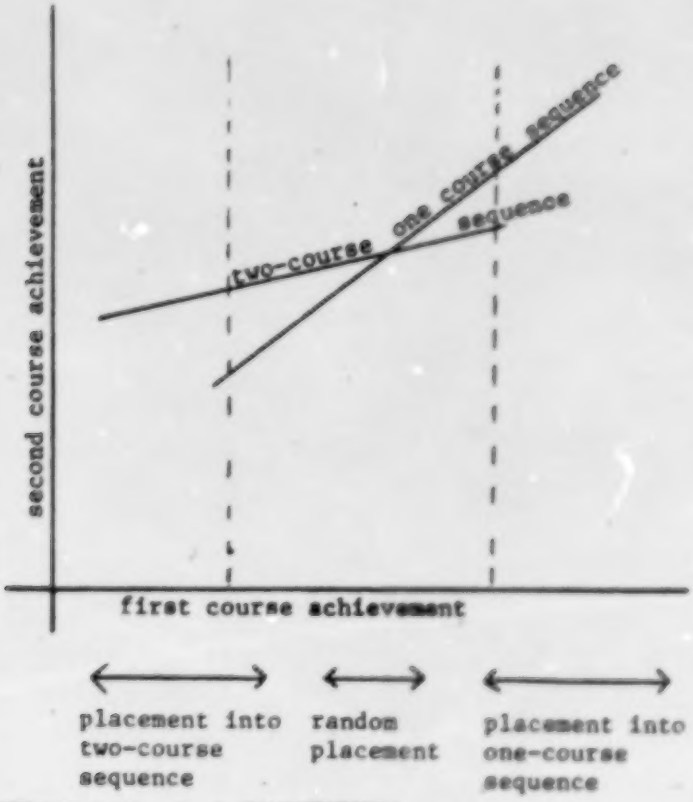
PROPOSED MODELS FOR EVALUATING CUTOFF SCORES USING TTI

We have found it impossible to evaluate the various methods for choosing cutoff scores in terms of concluding one method is better than another. This idea presupposes that there is a "real" value we wish to estimate. If there is a true cutoff score, it seems to us that it is the one determined by a trait-treatment interaction. Thus, we conclude this paper as we began, by proposing methods for evaluating cutoff scores for placement and exemption using the TTI framework. Further research on these methods is anticipated.

PLACEMENT

The problem is one of determining the regression lines for the two treatments under consideration. Theoretically, the way to do this would be to administer the trait assessment test, randomly assign students to the two treatments, and collect the criterion data. This procedure is not feasible in many real-life situations. One compromise on this method would be to randomly assign the students with scores in an indifference interval of the distribution of trait scores. Suppose that the problem under consideration is placement into first or second semester of a language course. Random placement of students with scores in such an interval, placing students with high scores into treatment A, and students with low scores into treatment B would yield the portions of the lines illustrated in Figure 9. Thus the intersection point of the two lines could be estimated if a linear assumption is made.

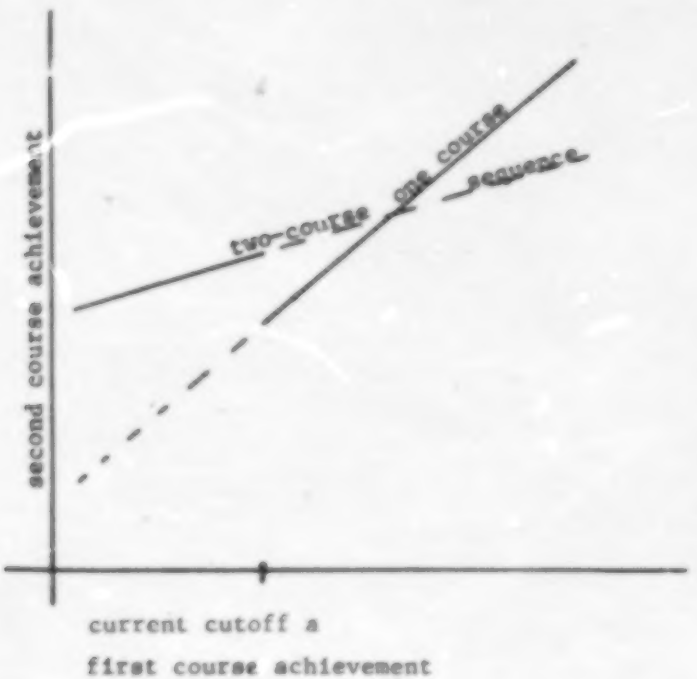
Figure 9: Random placement of students with trait scores in a specified interval.



The researcher sometimes does not even have the amount of control over placement that is demanded by the method of random assignment of students in an interval on the placement measures. Thus, we propose an alternative model for use in placement. This involves establishing an interim cutoff score and evaluating the results of using that cutoff. Hills (1971) has noted that once a cutoff score is established, seldom is it evaluated after use. Any of the methods described earlier might be used for establishing the interim cutoff score. Perhaps a cutoff score has been in use for some time, and now needs to be evaluated.

Three pieces of information are needed for each student: the placement test score, the course the student took (regardless of whether or not he/she took the recommended course), and the criterion score at the end of the second course in the sequence (or the "criterion" course in the case of remediation). The students may be self-selected if past data are used for this evaluation. If the placement test is not required of all students, then this may be precisely the group of students to whom we wish to generalize. (That is, the population to which the placement test is applied is a self-selected group, so using the self-selected group to evaluate the cutoff score would be the most representative sample.) Using the three pieces of information, regression lines for each treatment can be estimated. (See Figure 10.) The lines

Figure 10: Evaluation of a cutoff score that has been used in placement model.



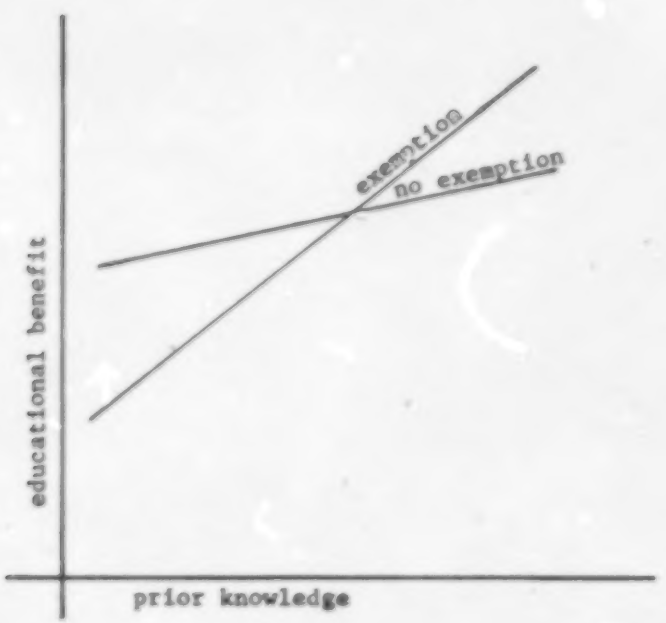
will not be estimated as accurately as they would in either of the previous two procedures, but it is unknown how unreliable this result would be. Further research on this method is also anticipated.

The placement models we have been considering involve placement into a long or a short sequence. Furthermore, two of the three models actually involve exemption from a prerequisite course. In these situations the exemption model may be judged to be more appropriate.

EXEMPTION

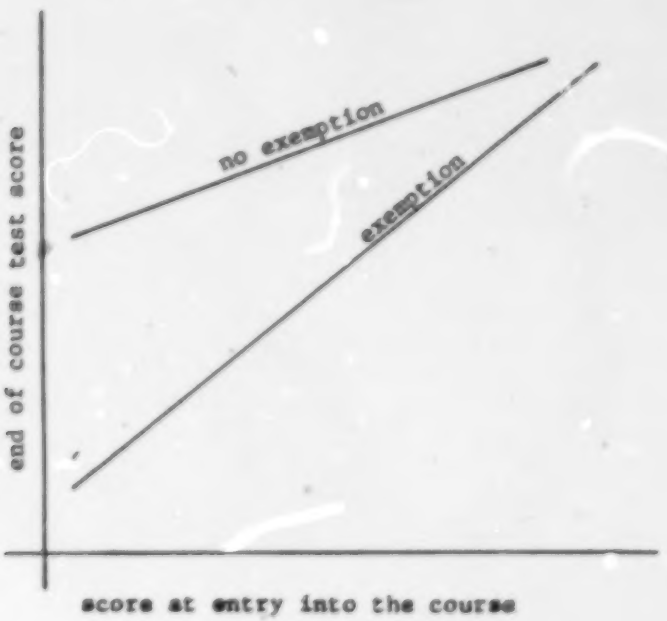
A difficulty encountered in applying TTI to exemption or credit by examination is that of determining a common criterion for measuring the educational outcome or educational benefit. TTI for exemption would take the form illustrated in Figure 11. To establish the regression

Figure 11: TTI for exemption or credit by examination.



lines for this situation, we propose administering one form of the test under consideration for exemption to students who are beginning the course. If the student is exempted from the course, his/her score at the beginning of the course might be thought to be identical with his/her score at the end of the course (this is subject to empirical verification). Thus, the regression line for exemption might be represented by regressing the 45° line on itself. The scores at the end of the course represent the outcome as a result of taking the course, so that the regression line for nonexemption is represented by the regression of end of course scores on entry into the course scores as illustrated in Figure 12. In translating this to educational benefit or benefit to the student, the line for exemption may be raised to

Figure 12: A proposed method for establishing TTI for exemption studies.



represent the saving time and the benefit that would be obtained from taking another course. Thus the graph in Figure 11 may represent a more realistic picture. The cutoff score for awarding credit or exemption would be based on the intersection point of the two lines. Further research on this procedure is also anticipated.

Figure 13 illustrates the application of the proposed method to the CLEP Introductory Sociology data. The exemption line is the regressed estimate obtained from entry scores (reliability = .85, mean = 47.0), and the nonexemption line is the regression of end on entry. The result is the ordinal interaction. (In this case, the interaction point for the regression lines falls outside the range of possible scores (20 to 80) or very nearly so. The average gain in scaled score units from entry to end was 11.4 (more than one standard deviation of the national sample of students in the norming group). If it can be assumed (subject to verification) that a student's expected gain from taking a different course is 11.4 units, then the line for exemption needs to be raised 11.4 units in translating the criterion variable from end of course test score to educational benefit. After this translation a disordinal interaction results as in Figure 14. The assumption that the slope of the exemption line does not change and is simply translated seems too strong. Persons at the upper end of the scale at entry will not tend to gain the same as persons at the lower end. Thus a relaxation of this assumption which would adjust the slope would probably be a more realistic approach. Further research on this method is anticipated.

Figure 13: Applying the proposed method for establishing TTI to CLEP Introductory Sociology data.

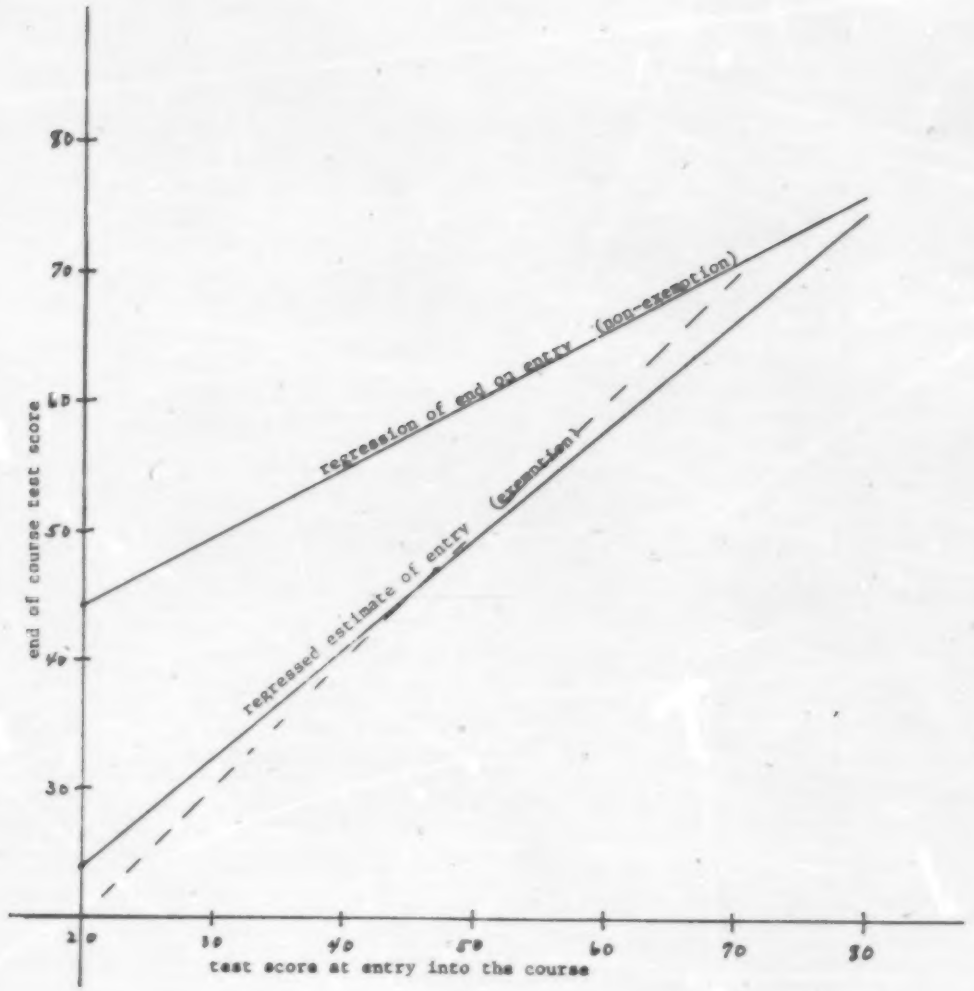
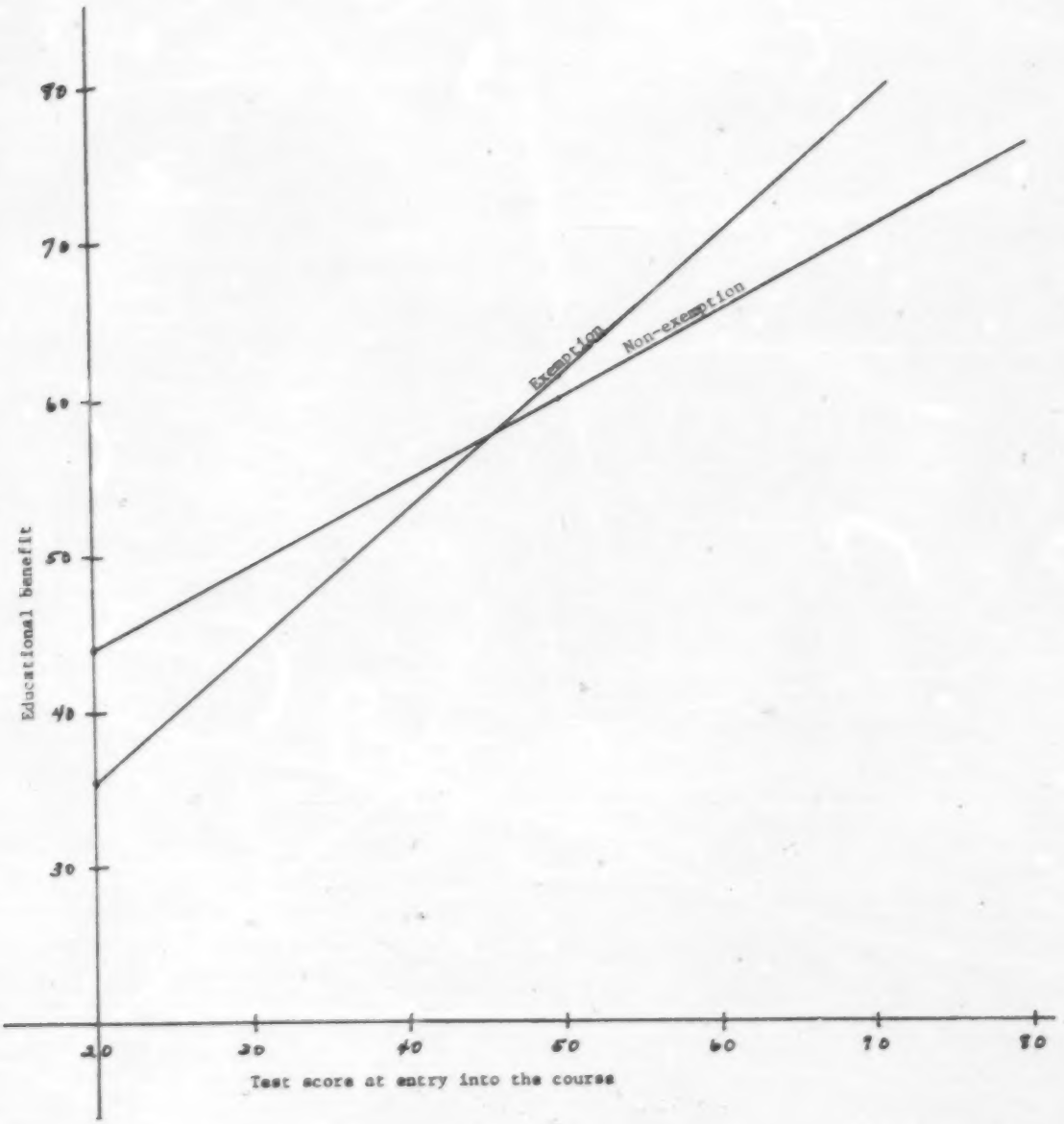


Figure 14: Applying the proposed method for establishing TTI to CLEP Introductory Sociology data after adjustment to educational benefit.



EPILOGUE

In this paper we have made many assumptions, offered predictions about what we expect to occur under certain conditions, and have applied some techniques which have not been thoroughly investigated. Most of these assumptions, predictions, and techniques are in elementary stages of verification at our institution, and we are interested in learning of the methods used at other institutions. We hope that this presentation of our experiences will facilitate communication of placement and exemption testing research.

REFERENCES

Aleamoni, Lawrence M. A study of foreign language learning at the University of Illinois. Measurement and Evaluation in Guidance, 1973, vol. 5, no. 4, pp. 468-474.

Berliner, David C., & Cahen, Leonard S. "Trait-Treatment Interaction and Learning" in Kerlinger, Fred N. (Ed.) Review of Research in Education. Itasca, Ill.: F. E. Peacock Publishers, INc., 1973, pp. 58-94.

Cohen, Allan S. Comparison of Decision-Making Strategies for Awarding College Credit Based on College Level Examination Program Examinations. (Doctoral dissertation, University of Iowa, 1972.) Ann Arbor, Michigan: University Microfilms, 1973, No. 73-13, 513. (Dissertation Abstracts Vol. 33, p. 6723-A.)

Cronbach, Lee J. "How can instruction be adapted to individual differences?" In R. M. Gagné (Ed.) Learning and Individual Differences. Columbus, Ohio: Charles E. Merrill, 1967, pp. 23-43.

Cronbach, Lee J. Beyond the Two Disciplines of Scientific Psychology. American Psychologist, 1975, vol. 30, no. 2, pp. 116-127.

Cronbach, Lee J., & Gleser, Goldine C. Psychological Tests and Personnel Decisions. Urbana, Ill.: University of Illinois Press, 1957. (rev. ed. 1965).

Flangher, Ronald L., Mahoney, Margaret H., & Messing, Rita B. Credit by Examination for College-Level Studies: An Annotated Bibliography. New York: College Entrance Examination Board, 1967.

Hills, John R. Use of Measurement in selection and placement, in Robert L. Thorndike (Ed.) Educational Measurement. (2nd ed.) Washington, D.C.: American Council on Education, 1971, pp. 680-732.

Klein, Anna C. CLEP German: A descriptive study (Placement and Proficiency Testing Report No. 5). Iowa City: Evaluation & Examination Service, February, 1977.

Klein, Anna C., & Whitney, Douglas R. CLEP Educational Psychology: Credit by examination validation study (Placement and Proficiency Report No. 1). Iowa City: Evaluation & Examination Service, October 1976.

Klein, Anna C., & Whitney, Douglas R. CLEP Introductory Sociology: Validation study (Placement and Proficiency Testing Report No. 2). Iowa City: Evaluation & Examination Service, December, 1976.

Kreplin, Hannah. Credit by Examination: A review and analysis of the literature. Berkeley, Calif.: University of California, 1971, 86 pp. ERIC ED 064 616.

Livingston, Samuel A. A utility based approach to the evaluation of pass-fail testing. ETS bulletin Copa-75-01, July, 1975. Princeton, New Jersey: Educational Testing Service.

Meskauskas, John A. Evaluation models for criterion-referenced testing: Views regarding mastery and standard-setting. Review of Educational Research, 1976, Vol. 46, pp. 133-158.

Perrin, David W. A comparison of methods for smoothing expectancy tables as applied to the prediction of success in college (Doctoral dissertation, University of Iowa, 1974). Ann Arbor, Michigan: University Microfilms, 1975, No. 75-13805. (Dissertation Abstracts Vol. 35, p. 7727-A.)

Perrin, David W. & Whitney, Douglas R. Methods for smoothing expectancy tables applied to the prediction of success in college. Journal of Educational Measurement, 1976, Vol. 13, pp. 223-231.

Tobias, Sigmund. Achievement Treatment Interactions. Review of Educational Research, 1976, Vol. 46, No. 1, pp. 61-74.

Webb, Sam C. Problems of placement programs. Paper presented at the regional meeting of the College Entrance Examination Board, New Orleans, Louisiana, February 23, 1967, 9 pp. as referenced in Willingham, 1974.

Willingham, Warren W. College Placement and Exemption. New York: College Entrance Examination Board, 1974.